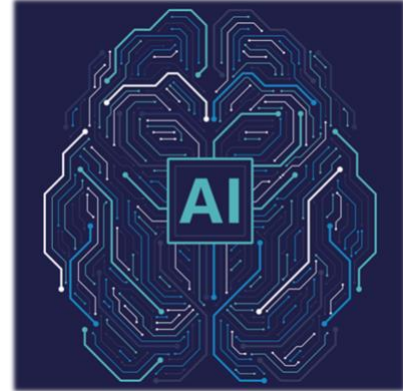


Generative AI Literacy of Incoming American College Students: Assessing the Need for Formal AI Literacy Training at Two Institutions



Authors:

Sara J. Finney, Ph.D.
James Madison University

Juste Mehou, M.A.
James Madison University

Stuart A. Miller, M.A.
Auburn University

Rachel Whitman Rotch, Ph.D.
Auburn University

Jaime Miller, Ph.D.
Auburn University

ABSTRACT

As artificial intelligence (AI) becomes embedded in education and the workforce, higher education faces growing pressure to ensure students are AI-literate, which necessitates high-quality measures of AI literacy and baseline assessment of AI literacy. We investigated the generative AI literacy of incoming college students from James Madison University (JMU) (n = 3037) and Auburn University (n = 3709) to (1) extend the evaluation of the Generative AI Literacy Assessment Test (GLAT), (2) assess baseline generative AI literacy prior to formal university training, and (3) explore links between generative AI literacy and student characteristics. The GLAT had adequate psychometric properties supporting its use. AI literacy was low for both institutions (63% versus 54% correct for Auburn and JMU, respectively) with negligible differences across student groups (e.g., major), underscoring the need for broad educational programming. Results added to literature on students' AI literacy and offered a method to assess curriculum needs for 21st-century competencies.

C V21_I3_RPA_2026_Generative_AI_Literacy_College_Students

Keywords: *generative AI, AI literacy assessment, needs assessment, higher education*

Call for AI Literacy in Higher Education

The rapid integration of artificial intelligence (AI) into nearly every facet of modern life—including healthcare, communication, and creative endeavors—has created an urgent need for AI literacy among students and professionals (Lintner, 2024). As AI technologies increasingly shape how individuals learn, work, and interact, higher education institutions are being called upon to prepare students not only to use AI tools but to understand, evaluate, and engage with them critically and ethically.

The urgency of this call is underscored by the growing presence of AI in the workplace. Employers increasingly demand AI-related skills, and job postings requiring AI competencies have surged in recent years. This trend is not limited to IT-related industries; AI is now embedded in careers as diverse as agriculture, architecture, health care, public administration, and management (Acemoglu et al., 2022; Alekseeva et al., 2021). As a result, students across all academic domains must be equipped with the knowledge and skills to navigate AI-enhanced environments (Mansoor et al., 2024; Verma et al., 2022).

Moreover, the proliferation of generative AI (i.e., “systems that can autonomously produce text, visuals, or other forms of media”; Jin et al., 2025, p. 1) introduced new challenges and opportunities for teaching and learning. These tools (e.g., ChatGPT) can enhance writing, support problem-solving, and foster creativity, but they also raise concerns about academic integrity, critical thinking, and overreliance on automated outputs. Thus, AI literacy must include not only technical proficiency but also the ability to use AI tools responsibly within specific disciplinary contexts (O’Dea & Ng, 2024; Sergeeva et al., 2025; Song, 2024; Tseng et al., 2025).

Put simply, AI literacy is now considered a foundational competency for the 21st century, akin to digital literacy (Southworth et al., 2023). AI literacy encompasses a broad set of skills, including *understanding* how AI systems work, *applying* AI tools in context, *evaluating* AI applications and outputs, and understanding *ethical* implications related to the use of AI (Ng et al., 2021a, 2021b; Ng et al., 2024a). These competencies are essential for fostering informed and responsible citizens in an AI-driven society. As noted by Laupichler et al. (2023, p. 1), AI literacy encompasses “competencies that include basic knowledge and analytical evaluation of AI, as well as critical use of AI applications by non-experts.” These are AI competencies that everyone should possess, including those without a computer science background (Ng et al., 2021b; Wang et al., 2023). Without AI competencies, people risk becoming passive consumers of AI rather than critical users.

To address this need, researchers and educators have emphasized the importance of measuring AI literacy accurately and continually (Laupichler et al., 2023; Mansoor et al., 2024; Wang et al., 2023). Reliable measures are essential for 1) identifying gaps in knowledge and need for education, 2) guiding AI literacy curriculum development, and 3) evaluating the effectiveness of educational interventions. Yet, a systematic review of AI literacy scales revealed that although some provide scores that are reliable and valid, few have been examined for cross-institutional or cross-cultural applicability or responsiveness to change (Lintner, 2024). Lack of robust measures has hampered efforts to track progress in AI literacy and design evidence-based educational strategies. More specifically, a critique of meta-analyses on AI in education revealed a key concern that weakened claims—selection of outcome measures (Bardach et al., 2025).

In sum, the current call for AI literacy in higher education is both timely and necessary. As AI continues to reshape society, the ability to understand, evaluate, and ethically engage with AI will be essential for academic success, career readiness, and civic participation. To meet this challenge, institutions must understand the needs of incoming students for timely educational programming to enhance AI literacy, which necessitates accurate assessment of AI literacy.

Measures Designed to Assess AI-Related Constructs

AI-related assessments can be broadly categorized into three types: measures of knowledge or competency, affective measures of attitudes or perceptions, and measures of use. Knowledge or competency measures aim to assess what individuals know or can do. These are cognitive tests or performance assessments that evaluate understanding of AI concepts (Hornberger et al., 2023; Jin et al., 2025). In contrast, attitudinal or perceptual measures focus on how individuals feel about AI (e.g., confidence), typically using Likert-scale self-report items (e.g., AI Self-efficacy Scale). Measures of use assess how frequently or in what ways individuals engage with AI, such as using ChatGPT or participating in AI-related learning activities (Pinski & Benlian, 2023). According to Lintner's (2024) systematic review, the most prevalent type of AI literacy measure is self-reported literacy (e.g., Meta AI Literacy Scale), which blends perceived knowledge with attitudes or usage patterns. These self-report tools may not accurately reflect competencies, highlighting the need for more robust assessments of AI literacy.

The Generative AI Literacy Assessment Test (GLAT; Jin et al., 2025) addresses this gap by offering a cognitive test designed to assess generative AI literacy. Similar to earlier tools such as Hornberger et al.'s (2023) AI literacy test, the GLAT uses multiple-choice questions; however, it differs in scope. The GLAT targets competencies specific to generative AI, which can independently produce text or visuals (Jin et al., 2025). The GLAT items represent four dimensions aligned with current conceptualizations of AI literacy by Ng and colleagues (Ng et al., 2021a, 2021b; Ng et al., 2024a): knowing and understanding (i.e., knowledge of concepts related to generative AI, zero-shot learning, prompt-development, content generation), using and applying (i.e., contextual understanding in content creation, information retrieval and synthesis, multimodal content generation, knowledge updating, token management), evaluating and creating (i.e., understanding of trustworthiness of output, hallucinations, content authenticity, knowledge cutoff), and ethical skills (i.e., understanding of responsible and ethical use of AI including privacy concerns, content safety, model bias, copyright issues). The GLAT is one of the few generative AI literacy tests evaluated through psychometric modeling and relations with external performance outcomes (i.e., GLAT scores predicted success in generative AI-supported tasks). This validity evidence positions the GLAT as a valuable tool for faculty and researchers seeking to assess generative AI competencies.

Levels of AI Literacy of College Students

Empirical research on college students' functional AI competencies is in the beginning stages. Early efforts to assess AI literacy across diverse college populations were conducted by Kong et al. (2021) and Kong et al. (2022) at the University of Hong Kong. In both studies, approximately 75% of participants were enrolled in undergraduate programs with the remainder in postgraduate or higher diploma programs. The students

represented a wide range of academic disciplines, including Chinese language studies, mathematics, English, and psychology. Prior to a seven-hour course on AI concepts delivered over three weeks, students completed a 14-item multiple-choice AI knowledge test assessing four specific skills: conceptual understanding, application, analysis, and evaluation. In both studies, pretest scores were around 45% correct. Post-test scores improved following the intervention (around 61% correct). Results highlighted two key findings. First, baseline AI literacy among university students was low. Second, AI literacy was amenable to improvement through educational intervention, even if improvement was modest. With that said, the psychometric properties of the test were suboptimal. The test lacked rigorous validation and the reliability of 0.66 fell below accepted standards.

To address these limitations, Hornberger et al. (2023) developed a more psychometrically sound multiple-choice test to measure general AI literacy. This test was administered to over 1,200 students from diverse academic disciplines (e.g., engineering, economics, social sciences, humanities) across six German universities. Students scored approximately 63% correct. Notably, students with prior AI coursework or a background in computer science and engineering significantly outperformed peers from non-technical fields such as the social sciences.

In a follow-up study, Hornberger et al. (2025) administered the same instrument to a multinational sample of more than 1,500 university students from Germany, the United Kingdom, and the United States. The majority of participants were pursuing a bachelor's degree (68.9%), whereas 26.8% were enrolled in master's programs and 4.2% in doctoral or other advanced programs. Across the full sample, students answered roughly 50% correct, lower than the average observed in the 2023 study of German students. U.S. students scored the lowest of the three groups. Interestingly, despite their lower performance, U.S. students reported higher AI self-efficacy, suggesting a misalignment between actual and perceived AI competence.

Collectively, these studies suggest that AI literacy is low for college students across several nations. Yet, no study has provided a detailed picture specific to U.S. students. Also, the measures used in these studies assessed general AI literacy, such as basic machine learning concepts, rather than focusing on generative AI literacy. Generative AI literacy is immediately necessary upon starting college as it is widely accessible and often assumed understood. Also, the extent to which AI literacy varies across subgroups of students (e.g., major) should be explored as it may uncover areas of need. These gaps in the literature underscore the need for this study, which compares students' generative AI literacy across two U.S. universities.

AI Literacy and Programming at Two U.S. Institutions

The two institutions involved in the current study had no indication of their students' AI literacy prior to this work. Moreover, there was limited discussion regarding institution-level AI literacy learning outcomes, coursework, or assessment prior to this study. Both challenges directly informed the purpose of this study, as discussed below.

AI Literacy at James Madison University

At James Madison University (JMU), there were no formal institutional-level student learning outcomes related to AI literacy or required educational training (e.g., courses, modules) to enhance AI literacy. Instead, the university has a formal institutional-

level learning outcome related to information literacy. Information literacy is defined by the faculty as students' ability to recognize the components of scholarly work and that scholarship can take many forms; demonstrate persistence and employ multiple strategies in research and discovery; identify gaps in their knowledge and formulate appropriate questions for investigation in academic settings; evaluate the quality of information and acknowledge expertise; use information effectively in their own work and make contextually appropriate choices for sharing their scholarship; and use information legally and ethically. Approximately every six years, first-year students complete an information literacy test designed by university faculty. To evaluate learning, students are re-assessed after completing information literacy educational programming and again as seniors.

Currently, the institution is engaged in a general education renewal process through which both student learning outcomes and educational programming are expected to be revised. Conversations swirled regarding the need for, type of, and placement of formal AI educational programming. A needs assessment was deemed essential to evaluate incoming students' AI literacy and to develop AI literacy programming. Not limited to general education, this needs assessment could inform broad institutional decision-making around AI-related teaching and learning—potentially including, but not confined to, the general education renewal process.

AI Literacy at Auburn University

Similar to JMU, Auburn had no formal institutional-level student learning outcomes related to AI literacy, nor required educational training to enhance AI literacy among students. The university relies on a student learning outcome centered around information literacy. This outcome was last assessed across both seniors and first-year students in 2021. To evaluate learning gains, a sample of first-year students is being re-assessed as seniors prior to graduation during the 2025-26 academic year.

Similar to JMU, Auburn is engaged in a general education renewal process through which both learning outcomes and educational programming are expected to be revised. Although there is currently no indication as to whether AI literacy will be formalized within the core outcomes, interest in integrating this technology into classrooms has sparked several independent initiatives across campus: faculty development courses, voluntary staff training, and AI summits hosted by industry leaders at the institution's teaching and learning center. Also, the AI Education Committee was formed to educate faculty, staff, and students about AI topics and support program design and implementation for AI literacy. Outreach efforts included the creation of AI integration courses for the broader community. However, there has not been an attempt to assess the impact of these efforts on AI literacy. Moreover, understanding incoming students' AI literacy is crucial for effective planning. Establishing a baseline will guide targeted efforts to improve AI literacy and inform inferences of their effectiveness.

Purpose of the Study

Although AI literacy is considered a foundational competency for the 21st century, many institutions are unaware of their students' AI literacy and, in turn, the need for intentional instruction. Thus, we had a three-fold purpose for the current study. First, we extended the evaluation of the psychometric properties of the GLAT. Second, we assessed incoming college students' levels of generative AI literacy at two institutions. Third, we

explored relations between generative AI literacy and students' characteristics (e.g., major), AI meta-cognition skills, agentic AI knowledge, and opinions of AI use and academic integrity.

Methods

Measures

The Generative AI Literacy Assessment Test (GLAT)

The GLAT was developed to assess competencies in using generative AI tools (Jin et al., 2025). The development process followed rigorous measurement standards, beginning with a blueprint of 25 generative AI concepts derived from academic literature, industry reports, and educational resources that were mapped onto the four dimensions of AI literacy (Ng et al., 2021a, 2021b; Ng et al., 2024a): knowing and understanding, using and applying, evaluating and creating, and ethics. Generative AI researchers reviewed the concepts and their mapping to ensure they covered the four dimensions of AI literacy. Multiple-choice items were created to align with generative AI concepts and both the items and distractors were reviewed by experts in generative AI, psychometrics, and educational psychology to ensure content validity. Ultimately, a 25-item test was created (see Appendix).

Five items were deleted from the GLAT during the validation process because they did not meet the developers' criteria for item discrimination (see items in Appendix with no psychometric properties). Discrimination indices reflect how well an item differentiates between individuals with varying levels of the underlying trait—in this case, generative AI literacy. Specifically, these items had discrimination indices below a threshold of 0.3 suggesting they were not effective at distinguishing between higher and lower levels of literacy (Jin et al., 2025).

Thus, the final version of the GLAT that the developers recommended is 20 items, each aligned with one of four dimensions of AI literacy. A total score is recommended rather than subscale scores as the test scores were deemed unidimensional via confirmatory factor analysis (Jin et al., 2025). The 20-item GLAT demonstrated adequate reliability (Cronbach's $\alpha = 0.80$). Item response theory (IRT) analyses indicated the GLAT provides the most information about below-average literacy individuals. Its predictive validity was supported; GLAT scores predicted performance on generative AI tasks, distinguishing it from self-report measures.

Given the limited study of the GLAT, we administered and examined all 25 items to evaluate if the same five items had low discrimination in our large samples, which would support their removal. If so, institution comparisons would be conducted using the 20-item GLAT.

Metacognition and Generative AI

Three additional generative AI items were developed by student support staff with AI expertise at Auburn because they felt the GLAT lacked specific questions tied directly to metacognition, a fifth domain beyond the four represented by the GLAT (see Table 1). The goal of including these items was to provide insight into students' understanding related to when and how to use AI tools. Specifically, we assessed if students could do the following: 1) explain the reasons for informed choices about using or not using generative AI for critical reading, writing, research, and learning purposes; 2) monitor generative AI

use in relation to their learning goals, including recognizing when an overreliance of generative AI tools could prevent their learning; 3) monitor how generative AI affects their development as a writer and identify when generative AI use limits their growth and potential. This is the initial study of the items' psychometric properties.

Agentic AI

Three items were developed by Auburn staff with AI expertise to assess student knowledge of and familiarity with Agentic AI (see Table 1). Whereas “traditional” AI agents are dependent upon human intervention in that they execute predefined tasks and respond to input, agentic AI is a more independent, proactive, and adaptive form of artificial intelligence (Murugesan, 2025). The items assess students' understanding of the difference between these two types of AI models as well as their ability to identify a key challenge of this technology. Auburn's staff that advocated for the inclusion of these three items felt that Agentic AI would likely be a form of AI that students would encounter more regularly during their time in college (i.e., the next wave of AI to impact their life). This is the initial study of the items' properties.

Use of AI and Perceptions of Academic Integrity

Auburn developed items to assess students' opinions of AI use and the connection of AI to academic integrity. Item 1 asked students how often they use AI for academic purposes (“Never” to “Always”). Item 2 assessed perceived risks related to generative AI and academic work (“Not Concerned at All” to “Extremely Concerned”). Item 3 asked students to consider the impact of AI tools on their ability to learn (“Strongly Negative Impact” to “Strongly Positive Impact”). The items were not written to represent a unified construct but to provide initial insight into if GLAT scores related to AI use (Item 1), perceived risks (Item 2), or learning impact (Item 3).

Participants and Procedures

Auburn University

Each summer, incoming first-year students are asked to complete four modules within a course on the Learning Management System (LMS) Canvas. The office that administers the course collects baseline data related to the general education student learning outcomes or some type of assessment that could be used to provide the institution baseline knowledge. Students were administered the GLAT during Summer 2025 and completed the assessment prior to attending on-campus orientation. The GLAT was administered remotely and students completed the unproctored test in an hour or less. The course opened at the beginning of May and closed at the end of July. Students were told test results would be used to inform educational program improvements but would not impact their grades or university record.

With a total incoming undergraduate class of 7,438, the GLAT was completed by 5,511 students. Of those, 3,709 students provided research consent, were at least 18 years of age, and completed testing by the deadline. Of this sample, 56% identified as female and 85% identified as White, reflecting the institution's demographics. Although the pre-orientation course in which the GLAT was administered was limited to first-year incoming undergraduate students, 528 participants were classified under an “Alternate First-Year” admissions type, indicating an initial admission to the university under an alternate path

such as dual or concurrent enrollment, global exchange, or similar programs. The remaining 3,181 students were directly admitted as “First-Time, First-Year” students. Across all admission types, the majority (98%) were registered as full-time students. Approximately 43% of the sample declared a STEM major. Student majors were coded as STEM or non-STEM using the institution’s standard classification of science, technology, engineering, and mathematics programs.

James Madison University

Data were collected for program assessment and accountability purposes during a university-wide Assessment Day (Aday) from entering first-year students in Fall 2025. Students were required to complete a series of assessments to avoid a registration hold the following term. The assessments were completed within a two-hour testing period in a remote and unproctored environment. The GLAT was the second test completed in the testing session. Students were told test results would be used to inform educational program improvements but performance would not impact their university record or grades.

A random sample of all students required to participate in Aday completed the GLAT ($n = 4,425$), and of those, 3,037 students provided research consent, were at least 18 years of age, and completed testing by the deadline. Of this sample, 55% identified as Female and 71% identified as White, reflecting the institution’s demographics. Approximately 23% declared a STEM major and 99.18% were full time. Student majors were coded as STEM or non-STEM using the institution’s standard classification of science, technology, engineering, and mathematics programs. All students in the sample were “First-Time, First-Year” students.

Results

Psychometric Properties of Scores

GLAT

Our first purpose was to extend the psychometric study of the GLAT. Item-level scores were examined using Classical Test Theory (CTT) and Item Response Theory (IRT) for each institution and the combined sample (see Appendix). Using CTT, reliability as well as item-level discrimination and difficulty were computed. Using IRT, the information function as well as discrimination, difficulty, and guessing parameters were computed.

The 25 GLAT items varied in discrimination and difficulty in both JMU and Auburn samples as well as the combined sample. For the combined sample, CTT discrimination (i.e., CTT- D^1) ranged from .08 to .63 across the 25 items, indicating some items were weak at differentiating between higher- and lower-performing students. In CTT, discrimination values below .20 are generally considered poor (signal items for revision or removal), whereas items with values of .30 or above indicate well-functioning items (Crocker & Algina, 1986). CTT difficulty (i.e., CTT- p , proportion answering correct) ranged from .26 to .88, where smaller values indicate more difficult items (.26 = 26%

¹ The CTT discrimination index was calculated by subtracting the proportion of correct responses from examinees with total scores above the 73rd percentile from the proportion of correct responses from examinees with total scores below the 27th percentile (see Bandalos, 2018, p.124)

answering correctly) and larger values indicate easier items (.88 = 88% answering correctly). IRT parameters ranged from 0.47 to 3.84 for discrimination (“a”) and -2.52 to 5.69 for difficulty (“b”), which paralleled variability in CTT indices.

Recall, in the initial study of the 25-item GLAT, five items (6, 9, 14, 15, 20) were removed due to poor performance (Jin et al., 2025). Consistent with findings from the initial study, items 6 (CTT-D = .13, CTT-p = .41), 14 (CTT-D = .08, CTT-p = .26), and 20 (CTT-D = .21, CTT-p = .70) displayed low discrimination in the combined sample. IRT results reinforced these concerns. For example, Item 6 had very low discrimination ($a = .52$, logistic metric) and an extremely high difficulty ($b = 5.69$), indicating the item contributed very little information about AI literacy and only at ability levels far above the typical measurement range (-3 to 3). Similarly, Item 14 had a high difficulty and only moderate discrimination ($b = 4.88$, $a = .77$). Item 20 had very low discrimination ($a = .33$). Taken together, these results supported the decision, consistent with the initial study, to remove these items from the 25-item scale. In contrast with the original GLAT study, items 9 (CTT-D = .38, CTT-p = .36) and 15 (CTT-D = .36, CTT-p = .37) did not fall below the discrimination threshold in our sample. Furthermore, Item 9 demonstrated high IRT discrimination ($a = 3.39$) with moderate difficulty ($b = 1.07$) and Item 15 showed a similar pattern ($a = 4.62$; $b = 1.06$). These findings suggest the removal of these items in the initial study may be due to sample variation and thus we recommend further study of these items.

For the remaining 20 items, which constitute the recommended 20-item GLAT, three (Item 1, CTT-D = .19; Item 2, CTT-D = .21; Item 3, CTT-D = .29) had low discrimination in our sample. Moreover, Items 1 and 3 were very easy ($p = .89$ and $.85$, respectively) whereas Item 2 was difficult ($p = .27$). IRT parameters supported these distinctions, with Item 1 ($b = -2.49$, $a = 0.81$) and Item 3 ($b = -1.70$, $a = 1.00$) providing information at lower ability levels, whereas Item 2 ($b = 1.40$, $a = 3.81$) was both highly discriminating and targeted higher ability students. Thus, although their CTT discrimination indices were low, the IRT results indicated moderate to strong discrimination across the target ability range (-3 to 3), providing no evidence for removal.

To facilitate cross-study comparisons, we retained the 20 items identified in the original study. The 20-item GLAT had adequate reliability ($\alpha = .77$ for the combined sample, $\alpha = .75$ for JMU, $\alpha = .76$ for Auburn). The average proportion correct was lowest for Ethics items (.49) and highest for Know and Understand items (.66). The items were slightly easier for Auburn than JMU students.

Agentic AI

All three items of the Agentic AI scale had proportion correct values at or below .56, indicating the items were difficult (see Table 1). Item 1 was the most difficult across samples. The items were also very highly discriminating across samples. Reliability of the three-item Agentic AI scale for the combined sample was acceptable ($\alpha = .73$); thus, we examined relations between the Agentic AI scale and the GLAT.

Metacognition Items

The metacognition items were easier than the Agentic AI items: combined sample proportion correct was .75, .71, and .65 for Items 1, 2, and 3, respectively. The metacognition items had low reliability ($\alpha = .56$); thus, a total score was not computed. Relations between each meta-cognition item and GLAT scores were examined.

Table 1. Agentic and Metacognition Item Characteristics for the JMU, Auburn, and Combined Samples

Scale	Item	JMU (n = 3,037)		Auburn (n = 3,709)		Combined (n = 6,746)	
		p	D	p	D	p	D
	Why is it important for students to be able to explain how they have used AI in projects?	0.70	0.73	0.79	0.45	0.75	0.72
Meta-cognition	How can students best ensure that using AI supports their learning rather than hinders it?	0.64	0.86	0.78	0.47	0.71	0.83
	How can students ensure that AI use supports their development as writers?	0.62	0.76	0.66	0.69	0.65	0.78
	What makes an "agentic AI" different from other types of AI?	0.34	0.81	0.47	1.00	0.41	1.0
Agentic AI	Which of the following is a key challenge associated with agentic AI?	0.40	0.84	0.56	1.00	0.49	1.0
	How does agentic AI behave differently from regular AI programs?	0.37	0.88	0.53	1.00	0.46	1.0
		JMU		Auburn		Combined	
		Mean	SD	Mean	SD	Mean	SD
Agentic AI Total Scores		1.10	1.10	1.56	1.24	1.35	1.20

Note. Items scored as correct (1) or incorrect (0). p = proportion correct. D = CTT discrimination. Agentic AI total could range from 0 to 3 and all possible scores were observed at both universities.

Levels of AI Literacy Across Institutions

Our second purpose was to examine the levels of incoming college students' generative AI literacy. The average AI literacy score on the 20-item GLAT collapsing across institutions was 11.79 (SD = 4.04) out of 20, corresponding to 59% correct (see Table 2). At Auburn, the average score was 12.63 (SD = 3.89; 63%), whereas at JMU it was 10.78 (SD = 3.97; 53.9%), an institution difference of about two points.

Within JMU, non-STEM students scored 10.69 on average, whereas STEM students scored 11.06, only 0.4 points higher on average. At Auburn, non-STEM students scored 12.43 on average, whereas STEM students scored 12.88, a difference of only 0.45 points on average. A similar pattern appeared for gender. At JMU, female students scored only 0.55 points higher on average. At Auburn, female students scored only 0.6 points higher on average. Across racial groups, average GLAT scores differed only modestly. Most group differences were under 1.5 points. Given the small sample sizes for some racial groups, results should be interpreted with caution. Considering admission type at Auburn,

Table 2. GLAT Scores by Student Demographics and Agentic AI scores for the JMU, Auburn and Combined Samples

	JMU					Auburn					Combined				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
	3,037	10.78	3.97	1	20	3,709	12.63	3.89	1	20	6,746	11.79	4.04	1	20
Stem major															
No	2,327	10.69	3.98	1	20	2,114	12.43	4.03	1	20	4,441	11.52	4.09	1	20
Yes	710	11.06	3.95	2	20	1,595	12.88	3.70	2	20	2,305	12.32	3.87	2	20
Gender															
Female	1,655	11.02	3.95	1	20	2,090	12.89	3.83	2	20	3,745	12.06	3.99	1	20
Male	1,380	10.47	3.98	2	20	1,619	12.29	3.95	1	20	2,999	11.45	4.07	1	20
Race															
Native Indian/Alaska						12	12.58	4.14	6	20	12	12.58	4.14	6	20
Asian	143	10.67	4.03	2	20	95	12.54	3.8	2	20	238	11.42	4.04	2	20
Black/African American	128	9.867	3.97	2	20	90	11.27	4.33	2	19	218	10.44	4.17	2	20
Hispanics of any race	337	10.61	3.87	2	20	206	12.55	3.99	3	20	543	11.35	4.02	2	20
Pacific Islander						2	10.00	2.83	8	12	2	10.00	2.83	8	12
Nonresident Alien	16	9.438	4.23	3	19	6	14.83	5.12	7	20	22	10.91	5	3	20
Race unknown	59	12.95	3.91	4	22	13	12.15	3.16	5	17	72	13.14	3.92	4	22
Two or More Races	193	11.03	4.03	2	20	129	12.89	3.67	2	20	322	11.77	3.99	2	20
White	2,161	10.84	3.99	1	20	3,156	12.66	3.88	1	20	5,317	11.92	4.02	1	20
Enrollment Status															
Full-time	3,012	10.78	3.98	1	20	3,630	12.66	3.88	1	20	6,642	11.81	4.03	1	20
Part-time	25	10.32	4.74	3	19	79	11.22	4.38	3	20	104	11.00	4.46	3	20
Admission Type															
“First-Time, First-Year”						3,181	12.67	3.80	2	20					
Alternate First-Year						528	12.37	4.43	1	20					

Note. “First-Time, First-Year” admission refers to students directly admitted to Auburn following completion of high school. “Alternate First-Year” admission includes students admitted to Auburn as part of dual enrollment, concurrent enrollment, global exchange, or similar programs. All students who took the test at JMU were “First-Time, First-Year” students. Two students at JMU did not identify as Male or Female.

students admitted as “First-Time, First-Year” students scored very similarly to their alternate admission counterparts. Negligible differences were found across part-time and full-time students at both institutions.

In sum, across institutions and demographic groups, students scored well below the maximum on the 20-item GLAT, averaging only about 12 correct responses ($\approx 59\%$). Although small differences emerged by major, gender, and ethnicity, these gaps were generally under one point on the 20-point scale. AI literacy does not practically differ across student sub-populations at these two institutions. All sub-populations had low AI literacy.

Relations between GLAT and Other Measures

Our third purpose was to examine the relation between the GLAT scores and other AI related constructs. Specifically, we examined whether generative AI literacy varied with agentic AI knowledge, metacognition, or beliefs about AI use. Total GLAT scores correlated moderately and positively with total Agentic AI scores in the combined sample ($r = .43$), with a somewhat stronger relation for Auburn ($r = .47$) than JMU ($r = .31$). Higher generative AI scores were associated somewhat with higher agentic AI scores.

Because of low reliability of the metacognition items, we inspected item-level contrasts (see Table 3). Recall, metacognition items were scored correct or incorrect. Students who selected the correct option for each item scored substantially higher on the GLAT. In the combined sample, GLAT score differences were 3.8 points for Item 1, 3.9 for Item 2, and 2.4 for Item 3. Although these items did not function as a single scale representing a unidimensional construct of meta-cognition, consistent score differences across items suggest that items may be capturing distinct aspects of a students’ understanding of when and how to use AI tools. Results indicated that higher generative AI knowledge was associated with higher AI meta-cognition skills.

Recall, the use and academic integrity items were not meant to be summed. For all three items, mean responses were below the midpoint of the five-point scale (see Table 4). This indicates the absence of ceiling effects—students did not report extremely high AI use for academic purposes (Item 1), strong concern about potential academic dishonesty (Item 2), or highly positive perceptions of AI’s impact on learning (Item 3). Instead, responses were distributed around the middle of each scale.

Endorsement of the use and integrity items varied negligibly with generative AI literacy (see Table 3). For Item 1, students endorsing “*Always—I rely on generative AI tools regularly for academic purposes*” scored much lower on the GLAT (combined $M = 6.90$) than peers choosing other options (11.7 to 12.0), a gap of about five points that was consistent across both institutions. The average generative AI knowledge increased very slightly with higher concern about potential academic dishonesty (Item 2) and no clear mean-level trend emerged for perceptions of AI’s impact on learning (Item 3). Thus, for the combined sample, the correlations between the three items and GLAT scores were near zero (Item 1: $r = -.04$; Item 2: $r = .06$; Item 3: $r = -.02$).

In sum, AI literacy covaried meaningfully with agentic AI knowledge and AI metacognition. AI use and integrity items showed little relation with GLAT scores, aside from markedly lower GLAT scores for students reporting frequent reliance on AI for academic work.

Table 3. Relation between GLAT Scores and AI Metacognition, Use, and Academic Integrity Items

		AI Metacognition Items by GLAT Score								
		JMU			Auburn			Combined		
		Mean	SD	N	Mean	SD	N	Mean	SD	N
Item 1	1	11.85	3.74	2,119	13.41	3.56	2,912	12.76	3.71	5,031
	0	8.29	3.35	918	9.75	3.71	797	8.97	3.59	1,715
Item 2	1	12.07	3.62	1,936	13.48	3.51	2,875	12.91	3.62	4,811
	0	8.5	3.53	1,101	9.7	3.73	834	9.02	3.66	1,935
Item 3	1	11.66	4	1,896	13.41	3.69	2,465	12.65	3.93	4,361
	0	9.3	3.46	1,141	11.08	3.81	1,244	10.23	3.76	2,385
		Use and Academic Integrity Items by GLAT Score								
Item 1	1	10.95	3.9	631	12.4	3.75	693	11.71	3.88	1,324
	2	10.89	4.03	1,137	12.65	3.69	1,506	11.89	3.94	2,643
	3	10.78	3.92	1,084	13.05	3.96	1,257	12.00	4.10	2,341
	4	10.05	4.01	159	11.58	4.52	218	10.94	4.37	377
	5	6.04	2.05	26	7.54	3.58	35	6.90	3.09	61
Item 2	1	10.16	4.03	604	12.31	4.05	778	11.37	4.18	1,382
	2	10.79	3.93	806	12.67	3.91	951	11.81	4.03	1,757
	3	10.88	4.02	938	12.63	3.96	1,090	11.82	4.08	2,028
	4	11.11	3.73	460	12.80	3.59	630	12.09	3.74	1,090
	5	11.23	4.13	229	12.94	3.74	260	12.14	4.02	489
Item 3	1	10.45	4.21	256	12.50	3.75	219	11.40	4.13	475
	2	11.32	3.89	819	12.78	3.69	1,026	12.13	3.85	1,845
	3	10.56	4.01	1,127	12.43	3.93	1,290	11.56	4.08	2,417
	4	10.82	3.74	710	12.88	3.93	1,025	12.04	3.98	1,735
	5	9.51	4.46	125	11.72	4.65	149	10.72	4.68	274

Note. Meta-cognition items were scored correct (1) or incorrect (0). AI use and academic integrity items were answered using a five-point scale. Item 1: 1 = Never, I do not use generative AI tools for academic purposes to 5 = Always—I rely on generative AI tools regularly for academic purposes. Item 2: 1= Not concerned at all to 5= Extremely concerned. Item 3: 1 = Strongly negative impact to 5 = Strongly positive impact.

Table 4. Academic Use and Integrity Item Descriptive Statistics for the JMU, Auburn, and Combined Samples

Item	Stem	JMU		Auburn		Combined	
		Mean	SD	Mean	SD	Mean	SD
1	How often do you use generative AI tools for academic purposes? Response Scale: (1) Never - I do not use generative AI tools for academic purposes. (2) Rarely – I use generative AI tools occasionally, but very infrequently. (3) Sometimes – I use generative AI tools occasionally, but not regularly. (4) Often – I use generative AI tools frequently for academic purposes. (5) Always - I rely on generative AI tools regularly for academic purposes.	2.28	0.88	2.30	0.87	2.29	0.87
2	How concerned are you about the potential for academic dishonesty with using generative AI tools? Response Scale: (1) Not concerned at all (2) Slightly concerned (3) Moderately concerned (4) Very concerned (5) Extremely concerned	2.64	1.18	2.63	1.19	2.64	1.18
3	To what extent do you believe the use of generative AI tools affects your learning and skill development? Response Scale: (1) Strongly negative impact (2) Somewhat negative impact (3) Neither negative or positive impact (4) Somewhat positive impact (5) Strongly positive impact	2.88	1.00	2.96	0.97	2.92	0.98

Note. All five response options were selected for all three items across both institutions.

Discussion

We assessed incoming students' generative AI literacy at two institutions to evaluate the need for formal AI literacy education. Using the GLAT, we found low levels of AI literacy with average scores of 54% at JMU and 63% at Auburn. Differences across student groups (e.g., major) were minimal, which some readers may find not surprising given the sample consisted of incoming students matriculating into college. That is, for incoming students, declared major is more aspirational than experiential, making the absence of differences across majors somewhat expected. In short, the findings suggest that low AI literacy is a widespread issue rather than confined to some majors. Given students' intended major was not associated with AI literacy levels, faculty and staff should not assume incoming students' AI-related knowledge and skills based on their major of interest. Recognizing this, the findings underscored the need for broad, rather than targeted, educational programming to enhance generative AI literacy.

Our findings align with prior research indicating that college students have low AI literacy levels (Hornberger et al., 2025; Kong et al., 2021; Kong et al., 2022). These prior studies focused on general AI literacy rather than generative AI. We extended this work by providing a detailed picture of generative AI literacy among first-year U.S. college students that confirmed low literacy levels across institution type and student demographics.

Implications for Assessment Practitioners

Results support the GLAT's utility for institutional needs assessments and curriculum development. It demonstrated adequate reliability ($\alpha = .77$) and provided usable information across a range of ability levels. The test could be completed in under 30 minutes, making it feasible for large-scale administration. Its psychometric properties were consistent with those reported by Jin et al. (2025) and we provided additional concurrent validity evidence via relations with agentic AI knowledge and metacognitive awareness.

Part of the rationale for assessing incoming students' agentic AI knowledge was that emerging agentic AI is likely to be a form of AI that students would encounter more regularly during their time in college (i.e., the next wave of AI to impact their lives). These items were difficult for students (proportion correct $\leq .49$ in the combined sample) with high discrimination indices. Although the items would benefit from further psychometric investigation, they can be used in practice as a short assessment tool. Because the items discriminate strongly and the three-item reliability is acceptable ($\alpha = .73$), the composite is defensible for group-level decisions.

When examining the metacognition items further, those items showed relatively low internal consistency and it is likely that this reflects the characteristics of metacognition itself rather than a measurement issue. Metacognition related to generative AI use appears to include multiple skills that may or may not overlap, such as knowing when AI support is appropriate or necessary, evaluating the quality and limitations of an AI-generated outputs, and managing how much someone should truly rely on AI tools. Additionally, the three items highlight different metacognitive functions, including reflection and transparency around AI use (Item 1), self-regulation of AI in support of learning (Item 2), and domain-specific considerations related to writing development (Item 3). These skills could present unevenly across students and/or lack of consistency

across tasks, especially in settings where AI usage depends on what a student hopes to accomplish or achieve. Given that context, these items may be most informative when examined individually as key indicators of specific metacognitive skills. Viewed this way, lower reliability may indicate variation in how students consider their AI usage with some areas presenting stronger than others. In sum, treating AI metacognition as a set of distinct, diagnostic skills (rather than a single unified trait) would showcase how students understand and manage generative AI and offer useful direction for future assessment and teaching practices. We encourage assessment practitioners to look further into the measurement of AI metacognition.

Implications for AI Literacy Educational Programming

The consistently low scores across institutions and student groups indicate a need for early, formal, and broad generative AI training. Students struggled most with ethical and evaluative dimensions, suggesting that programming should emphasize responsible use, critical evaluation of AI outputs, and understanding privacy and bias. Moreover, students with lower AI literacy were more likely to overuse AI tools and perceive them as beneficial, which highlights the importance of embedding metacognitive training into AI literacy curricula.

At JMU, results were shared with upper administration. There has been no decision at this point to add a university-wide student learning outcome related to AI. The revised general education curriculum may incorporate AI-related outcomes, but it is too early to know. Moreover, it is unclear if and how academic degree programs are currently incorporating AI-related learning outcomes into their curricula. Regarding educational programming, the university's AI taskforce had a key recommendation regarding AI literacy. "Enhance AI Literacy for Faculty and Students: Integrate AI literacy into curricula to ensure that faculty and students understand how to use AI tools responsibly, ethically, and effectively. This is hoped to foster a deeper comprehension of why, when, and how AI can be applied in their work." However, the current approach to increasing AI literacy remains optional workshops or guides offered by the university library and information technology units.

At Auburn, results were shared with the Core Curriculum General Education Committee where members were engaging in a core curriculum review. The findings of the study resonated with a research librarian serving on the committee who indicated students were seeking more guidance on how to use AI for research papers. Currently, Auburn has no formal, centralized policy or guidance on how faculty should address AI usage. AI integration appears limited to select graduate programs, primarily within the College of Engineering. The decentralized nature of the institution leaves decisions largely up to individuals. Future discussions with upper-level academic leaders are expected to explore the study's findings and highlight the need for a clear, consistent, institution-wide policy regarding AI training. Until such a policy exists, students will encounter varied support across campus. To date, no policy changes have taken place at the upper administration level of the institution, making it difficult to track how academic degree programs are weaving AI usage into their curricula or learning outcomes. Regardless of whether a policy is implemented, findings from this study indicate that low AI literacy among students may contribute to AI misuse and lack of post-graduation preparation.

Call for Institutionalizing AI Literacy: A Campus-Wide Strategy

Despite growing urgency for AI literacy (Tzirides et al., 2024), higher education has struggled to respond. Although U.S. institutions account for 23% of global AI literacy research (Biagini, 2025), few have translated those findings into institutional practice (Laupichler et al., 2023). Further, faculty engagement with AI is notably higher than with staff or students (Daniels, 2025). Echoing others, AI literacy curricula should be coherently integrated and sequenced in and out of the classroom, which necessitates AI literacy for faculty and staff as discussed below.

In and Out of the Classroom: Pedagogy, Ethics, and Applied Problem-Solving

We argue that AI literacy must be embedded throughout curricular and co-curricular learning rather than relegated to one-off training modules. Faculty can design activities that explore complex, real-world problems and ethical dilemmas (Černý, 2024), require critical review and inquiry with AI concepts (Sari et al., 2025), employ varied roles of AI for class projects (e.g., mentor, coach, simulator; Mollick & Mollick, 2023), and adopt project-based approaches (Kong et al., 2025). Ng and colleagues (2024a) clarified the complementary roles of pedagogy, materials, and assessment to enhance AI literacy. Intentional instructional design, including experiential-based pedagogical choices (Sari et al., 2025) and gamification of AI concepts (Ng et al., 2024b), should improve students' application of AI. These strategies to increase AI literacy can be employed in beyond-the-classroom experiences including residence hall programming, first-year experience seminars, peer tutoring, and practicums.

Build Capacity: Faculty and Staff Development for All (Not Just the Eager Few)

Enhancing students' AI literacy requires educators who are themselves AI literate. Yet, even if materials for educators exist (e.g., Auburn's AI module-based course), the resources are often stored within teaching and learning faculty spaces and are optional engagements that lack accountability or recognition. Only the most passionate educators elect into training and the learnings are not required to be embedded in curricula or campus programming, meaning students can complete college without encountering them. Likewise, crowdsourced ideas for assignments and teaching approaches (Barnard College Center for Engaged Pedagogy, n.d.) and curated resources on using generative AI (Artificial Intelligence (AI) and Technology, n.d.; Auburn University - University Writing, n.d.) are helpful, yet these materials alone cannot meaningfully enhance AI literacy unless they are actively experienced.

Instead, institutions may consider models like University of Florida's (UF) "AI Across the Curriculum" (Southworth et al., 2023). This comprehensive initiative embeds AI education into general education, co-curricular, and disciplinary courses, making it a core priority. To do so, UF invested in new technology, faculty expertise, staff upskilling, and assessment. As a result, undergraduate students connect with AI literacy concepts through undergraduate research, career preparation, experiential learning, and certification programs. Beyond a single institution, EDUCAUSE has developed guidelines related to AI Literacy (Hibbert et al., 2024). Or as a starting point, consider Zhou and Schofield's (2024) conceptual model for how to integrate AI literacy into course instruction or Chen and colleagues (2025) eight-week *AI Academy* as a scalable model for

educator cohort-based learning. In short, institutions should take an initial step toward integrating AI literacy—doing nothing is no longer an option.

Limitations and Future Research

Researchers may want to consider the role of floor effects, low-stakes assessments, and data collection design for future studies of AI literacy. With respect to floor effects, when the majority of students are performing very poorly, lack of large differences between groups does not necessarily mean equivalent knowledge levels. The assessment may not be sensitive enough to capture meaningful variation in students' (very low) understanding, even if such variation exists. Future studies could administer the GLAT after targeted instructional interventions or later in students' academic careers to determine whether group differences exist as AI literacy increases.

The GLAT was a low-stakes assessment for the students completing it at both Auburn and JMU; performance on this test had no personal consequence for the student. Low-stakes assessments often have issues with examinee effort (e.g., Schaefer & Finney, 2025; Wise & Demars, 2010). Future studies could gather measures of examinee effort (e.g., Pastor et al., 2023), filter students' scores who expended low effort to provide more accurate estimates of AI literacy (e.g., Swerdzewski et al., 2011), and apply interventions to increase effort on these assessments (e.g., Finney et al., 2025).

Finally, the cross-sectional data collection design of this study cannot draw conclusions about how students' generative AI literacy develops over time or in response to instruction. Future longitudinal studies that follow students across their college experiences, particularly before and after AI-embedded coursework, would allow researchers to examine growth trajectories and better understand the differences among discipline-specific (e.g., majors) experiences.

In sum, although the present study provides a valuable snapshot of incoming students' generative AI literacy, its limitations underscore the importance of future research that tracks learning longitudinally and evaluates targeted instructional efforts. Research examining changes in AI literacy as students progress through discipline--specific coursework will be especially critical for identifying where and how institutions can intervene most effectively. By pursuing these lines of inquiry, universities will be better positioned to create coherent, campus-wide strategies that cultivate the ethical, evaluative, and applied AI competencies students need for academic and professional success.

Appendices

[Appendix: 20-item GLAT characteristics in the JMU, Auburn, and Combined Sample](#)

References

- Acemoglu, D., Autor, D., Hazell, J. & Restrepo, P. (2022). Artificial intelligence and jobs: evidence from online vacancies. *Journal of Labor Economics*, 40(S1), S293–S340. <https://doi.org/10.1086/718327>
- Alekseeva, L., Azar, J., Giné, M., Samila, S., & Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, 71, 102002. <https://doi.org/10.1016/j.labeco.2021.102002>
- Artificial Intelligence (AI) and Technology. (n.d.). DePaul University. <https://condor.depaul.edu/jmoore/tech/ai/>
- Auburn University - University Writing. (n.d.). *Resources: Artificial intelligence and writing*. University Writing. <https://auburn.edu/academic/provost/university-writing/resources/?tag=Artificial+Intelligence>
- Bardach, L., Emslander, V., Kasneci, E., Eitel, A., Lindner, M. A., & Bailey, D. H. (2025). *Research syntheses on AI in education offer limited educational insights* [Preprint]. OSF. https://doi.org/10.31219/osf.io/dx6kt_v1
- Barnard College Center for Engaged Pedagogy. (n.d.). *Integrating generative AI in teaching and learning*. <https://cep.barnard.edu/integrating-generative-ai-teaching-and-learning>
- Biagini, G. (2025). Towards an AI-literate future: A systematic literature review exploring education, ethics, and applications. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-025-00466-w>
- Černý, M. (2024). University students' conceptualisation of AI literacy: Theory and empirical evidence. *Social Sciences*, 13(3), 129. <https://doi.org/10.3390/socsci13030129>
- Chen, S., Tang, X., Cheng, A., Chawla, N., Ambrose, G. A., & Metoyer, R. (2025). *Teaching the Teachers: Building Generative AI Literacy in Higher Ed Instructors*. <https://arxiv.org/html/2509.11999v1>
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston.
- Daniels, C. P. (2025). *Exploring artificial intelligence literacy and engagement in higher education stakeholders* (Doctoral dissertation, Middle Georgia State University). https://comp.mga.edu/static/media/doctoralpapers/2025_Daniels_0715161106.pdf
- Finney, S. J., Miller, S. A. & McGoey, K. (2025). Increasing expended effort on low-stakes accountability tests via priming: Effectiveness with graduating university students. *Research & Practice in Assessment*, 20(1), 76–92. <https://files.eric.ed.gov/fulltext/EJ1488121.pdf>
- Hibbert, M., Altman, E., Shippen, T., & Wright, M. (2024, June 3). *A framework for AI literacy*. EDUCAUSE Review. <https://er.educause.edu/articles/2024/6/a-framework-for-ai-literacy>
- Hornberger, M., Bewersdorff, A., & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, 5, Article 100165. <https://doi.org/10.1016/j.caeai.2023.100165>
- Hornberger, M., Bewersdorff, A., Schiff, D. S., & Nerdel, C. (2025). A multinational assessment of AI literacy among university students in Germany, the UK, and the US. *Computers in Human Behavior: Artificial Humans*, 5, Article 100132. <https://doi.org/10.1016/j.chbah.2025.100132>
- Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The Generative AI Literacy Assessment Test. *Computers and Education: Artificial Intelligence*, 9, Article 100436. <https://doi.org/10.1016/j.caeai.2025.100436>

- Kong, J., Liu, J., Chen, G., & Shang, W. (2025). Assessing AI literacy in college students: The mediating role of self-efficacy in motivational commitment pathways. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-025-13753-9>
- Kong, S.-C., Cheung, W. M.-Y., & Tsang, O. (2022). Evaluating an artificial intelligence literacy programme for empowering and developing concepts, literacy and ethical awareness in senior secondary students. *Education and Information Technologies*, 28, 4703–4724. <https://doi.org/10.1007/s10639-022-11408-7>
- Kong, S.-C., Cheung, W. M.-Y., & Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence*, 2, Article 100026. <https://doi.org/10.1016/j.caeai.2021.100026>
- Laupichler, M. C., Aster, A., Haverkamp, N. & Raupach, T. (2023). Development of the “Scale for the assessment of non-experts’ AI literacy” – An exploratory factor analysis. *Computers in Human Behavior Reports*, 12. <https://doi.org/10.1016/j.chbr.2023.100338>.
- Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning*, 9(50). <https://doi.org/10.1038/s41539-024-00264-4>
- Mansoor, H. M. H., Bawazir, A., Alsabri, M. A., Alharbi, A. & Okela, A. H. (2024). Artificial intelligence literacy among university students—A comparative transnational survey. *Frontiers in Communication*, 9, Article 1478476. <https://doi.org/10.3389/fcomm.2024.1478476>
- Mollick, E. & Mollick, L. (2023). *Assigning AI: Seven Approaches for Students with Prompts*. The Wharton School Research Paper. <https://arxiv.org/pdf/2306.10052>
- Murugesan, S. (2025). The rise of agentic AI: implications, concerns, and the path forward. *IEEE Intelligent Systems*, 40(2), 8–14. <https://doi.org/10.1109/MIS.2025.3544940>
- Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021a). AI literacy: Definition, teaching, evaluation and ethical issues. *Proceedings of the Association for Information Science and Technology*, 58(1), 504–509. <https://doi.org/10.1002/prai.2.487>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021b). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Ng, D. T. K., Su, J., Leung, J. K. L., & Chu, S. K.W. (2024a). Artificial intelligence (AI) literacy education in secondary schools: A review. *Interactive Learning Environments*, 32(10), 6204–6224. <https://doi.org/10.1080/10494820.2023.2255228>
- Ng, D. T. K., Xinyu, C., Leung, J. K. L., & Chu, S. K. W. (2024b). Fostering students' AI literacy development through educational games: AI knowledge, affective and cognitive engagement. *Journal of Computer Assisted Learning*, 40(5), 2049–2064. <https://doi.org/10.1111/jcal.13009>
- O’Dea, X., & Ng, D. T. K. (2024). AI literacy and gen-AI Literacy frameworks. In X. O’Dea and D.T.K. Ng (Eds.), *Effective Practices in AI Literacy Education: Case Studies and Reflections* (pp. 21–27). Emerald Publishing Limited. <https://doi.org/10.1108/9781836088523>
- Pastor, D. A., Patterson, C. R., & Finney, S. J. (2023). Development and internal validity of the Student Opinion Scale: A measure of test-taking motivation. *Journal of Psychoeducational Assessment*, 41(2), 209–225. <https://doi.org/10.1177/07342829221140957>

- Pinski, M. & Benlian, A. (2023). AI literacy – Towards measuring human competency in artificial intelligence. *Proceedings of the 56th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2023.021>
- Sari, D. K., Supahar, S., Rosana, D., Dinata, P. A. C., & Istiqlal, M. (2025). Measuring artificial intelligence literacy: The perspective of Indonesian higher education students. *Journal of Pedagogical Research*, 9(2), 143–157. <https://doi.org/10.33902/JPR.202531879>
- Schaefer, K. E., & Finney, S. J. (2025). Who is testing at midnight? College students' disengagement throughout a 24-hour low-stakes testing window. *Educational Assessment*, 30(4), 249–264. <https://doi.org/10.1080/10627197.2025.2487108>
- Sergeeva, O. V., Zheltukhina, M. R., & Demir, S. (2025). From assistance to integrity: Exploring the role of AI in academic communication across Russian and Turkish campuses. *Frontiers in Communication*, 10. <https://doi.org/10.3389/fcomm.2025.1479813>
- Song, N. Y. (2024). Higher education crisis: Academic misconduct with generative AI. *Journal of Contingencies and Crisis Management*, 32, e12532. <https://doi.org/10.1111/1468-5973.12532>
- Southworth, J., Migliaccio, K., Glover, J., Glover, J., Reed, D., McCarty, C., Brendemuhl, J. & Thomas, A. (2023). Developing a model for “AI Across the Curriculum”: Transforming the higher education landscape via innovation in AI literacy. *Computers and Education: Artificial Intelligence*, 4, Article 100127. <https://doi.org/10.1016/j.caeai.2023.100127>
- Swerdzewski, P. J., Harmes, J. C., & Finney, S. J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education*, 24, 162–188. <https://doi.org/10.1080/08957347.2011.555217>
- Tseng, L.-P., Huang, L.-P., & Chen, W.-R. (2025). Exploring artificial intelligence literacy and the use of ChatGPT and copilot in instruction on nursing academic report writing. *Nurse Education Today*, 147, Article 106570. <https://doi.org/10.1016/j.nedt.2025.106570>
- Tzirides, A. O., Zapata, G., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., You, Y., Afonso dos Santos, T., Sears Smith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, 6, 100184. <https://doi.org/10.1016/j.caeo.2024.100184>
- Verma, A., Lamsal, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industrial and Higher Education*, 36(1), 63–73. <https://doi.org/10.1177/0950422221990990>
- Wang, B., Rau, P. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 42(9), 1324–1337. <https://doi.org/10.1080/0144929X.2022.2072768>
- Wise, S. L., & DeMars, C. E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, 15(1), 27–41. <https://doi.org/10.1080/10627191003673216>
- Zhou, X., & Schofield, L. (2024). Developing a conceptual framework for artificial intelligence (AI) literacy in higher education. *Journal of Learning Development in Higher Education*, 31. <https://doi.org/10.47408/jldhe.vi31.1354>